MARTA CATILLO

Consortium GARR | THE ITALIAN EDUCATION & RESEARCH NETWORK

A deep learning intrusion detection system

GIORNATA DI INCONTRO
BORSE DI STUDIO GARR
"ORIO CARLINI"
ROMA

Roma, 27 Giugno 2019

Borsisti Day 2019

Sintesi dell'attività

Obiettivo

Uso di tecniche di Deep-Learning per la detection di 0-day attacks e realizzazione di un network Intrusion Detection System anomaly-based

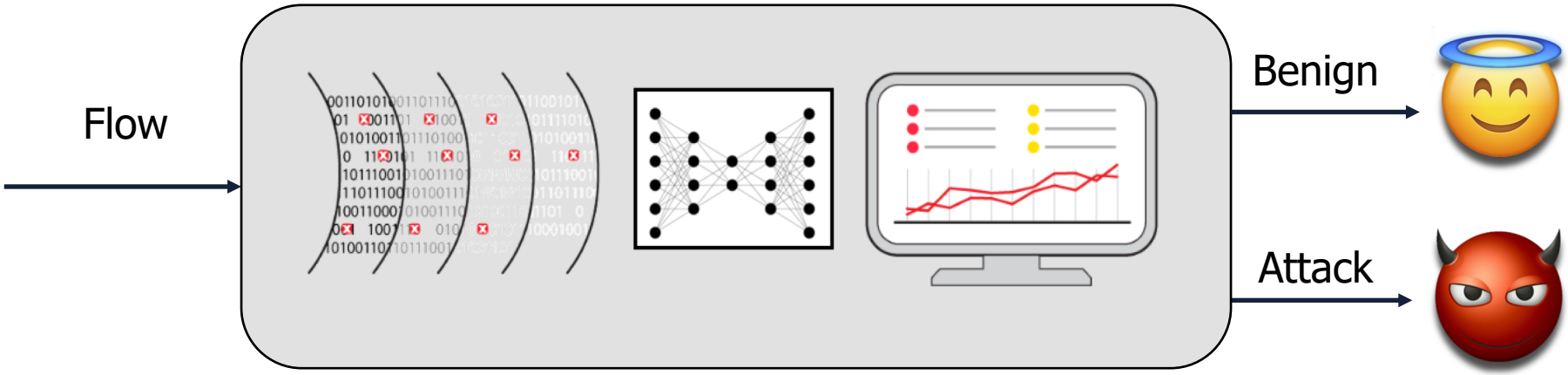
Motivazione

Gli attuali IDS sono inefficaci per attacchi non noti

Sede

Dipartimento di Ingegneria dell'Università degli Studi del Sannio-DING

Overview



- NIDS anomaly-based
- Basato sull'analisi di flussi
- Modello **Autoencoder**
  - Problema trattato come task **semi-supervised**

Strengths

Architettura 'snella'

Deep Autoencoder autonomo

Semi-supervised learning

- Processo di learning veloce
- Learning nell'ordine di minuti anche su computer desktop

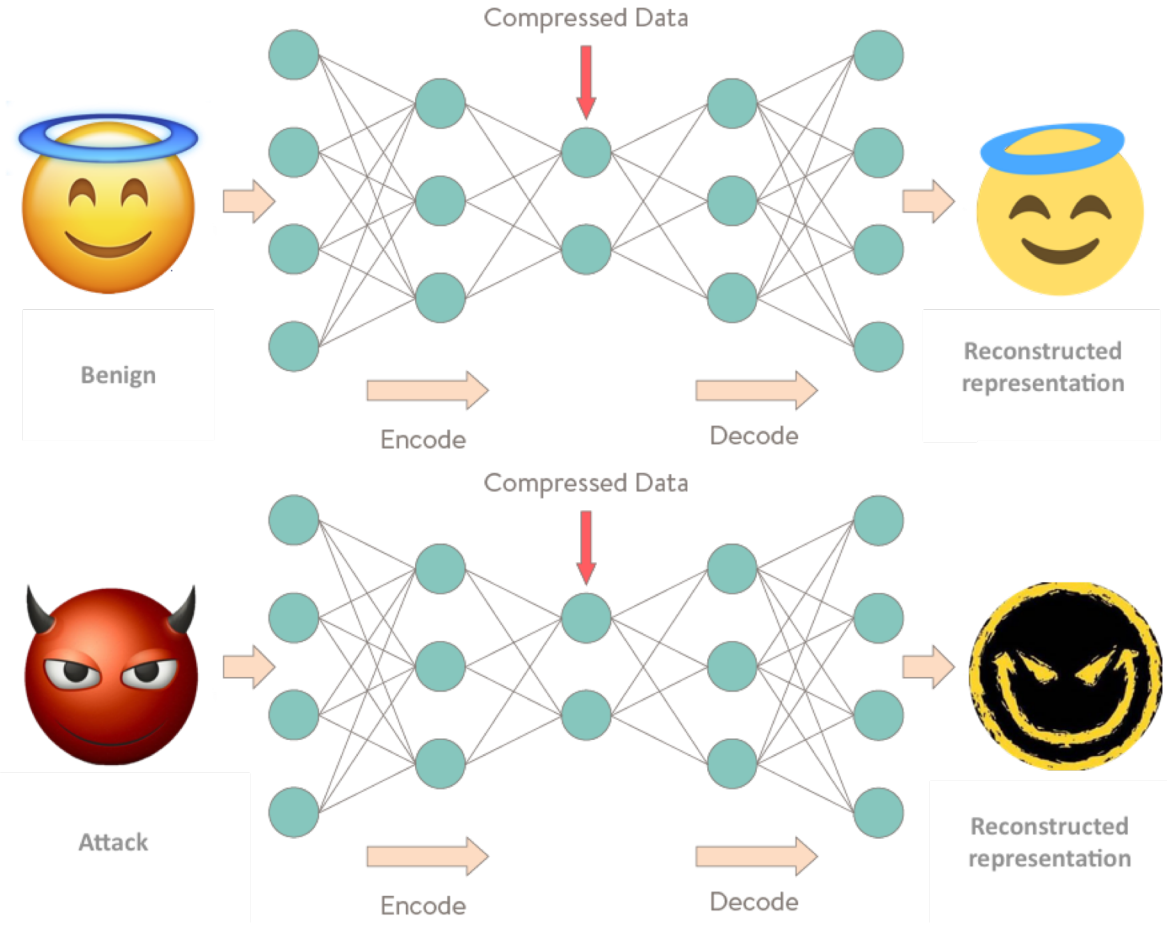
State-of-the-art dataset

Estesa sperimentazione sul recente dataset CICIDS2017

0-day discovery

Possibilità di scoprire attacchi non previsti nel training set

Metodologia (1)



Training e validazione con traffico Benigno

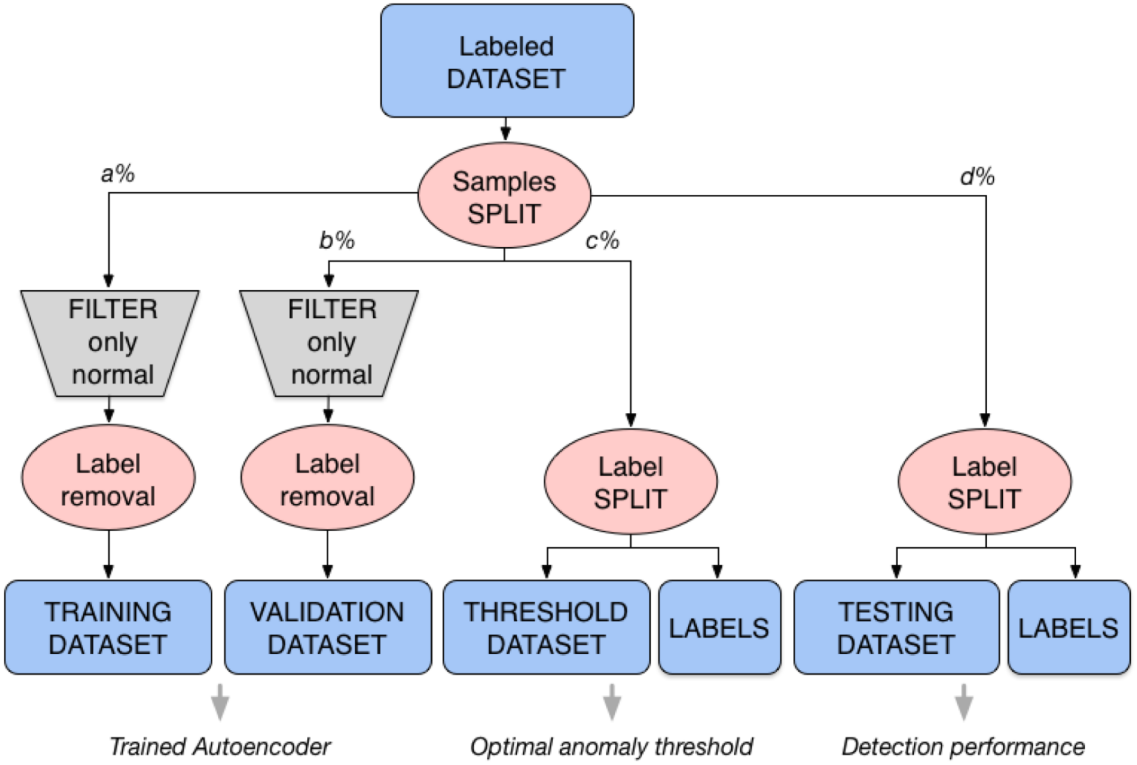
Low Reconstruction Error

High Reconstruction Error

ALERT!!!

…*threshold* RE setting

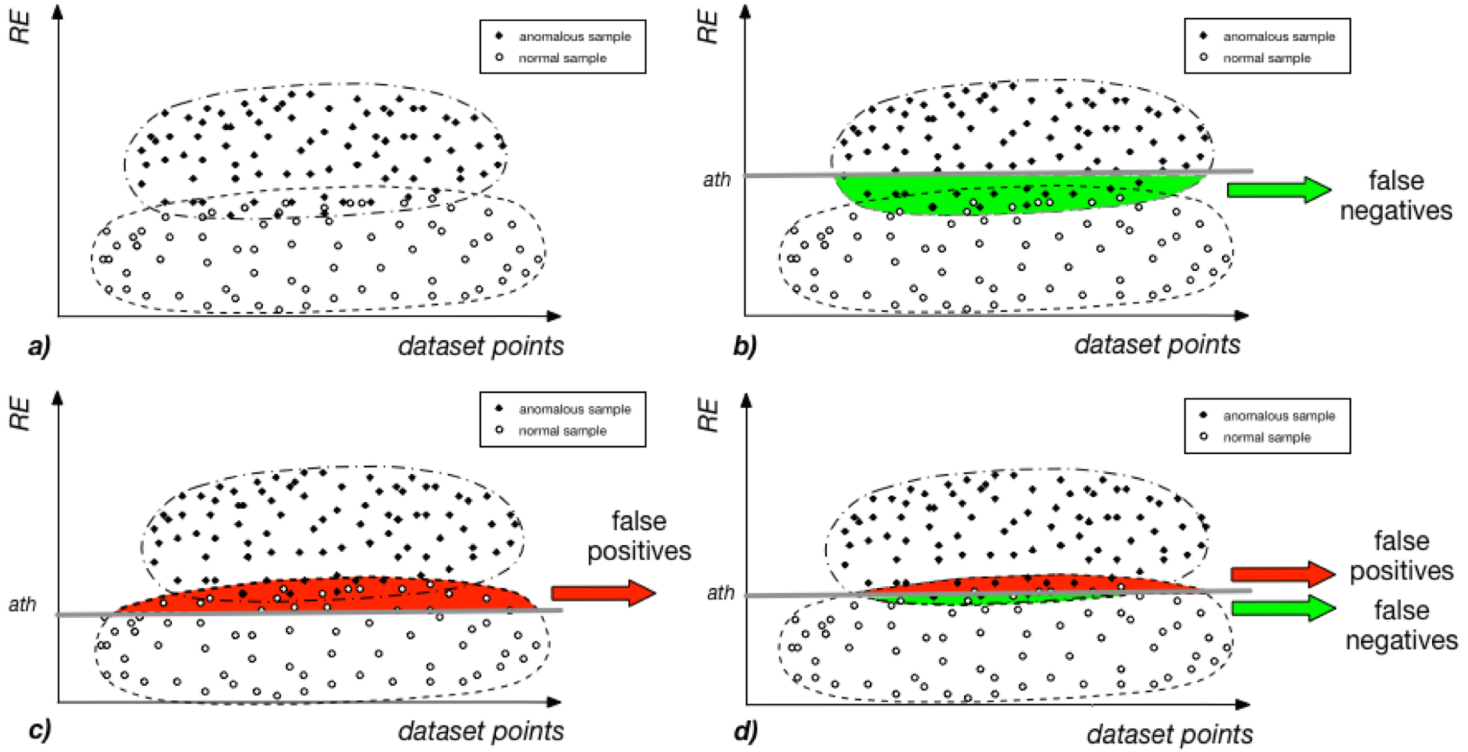
Metodologia (2)



Partizionamento del dataset labelizzato

- Output delle fasi di ***training*** e ***validazione***: **AE** capace di ricostruire traffico benigno con **RE** basso.
- **threshold dataset** e **testing dataset** non filtrati (traffico normale + attacchi) con separazione delle label.
- scelta di una ***threshold*** ottima.
- unlabeled sample del **threshold dataset** processati dall'**AE** addestrato.
- **RE** associato all'informazione contenuta nella label del sample (traffico normale o attacco).
- processing dei sample presenti nel **testing dataset**.

Metodologia (3)



Anomaly Threshold setting

- Trade-off tra ***precision*** e ***recall***: massimizzazione dell'***F1 score***

Sperimentazione (1)

Attacchi

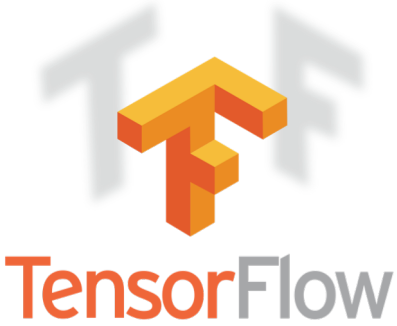
Class Labels	Number of instances
BENIGN	2359087
DoS Hulk	231072
PortScan	158930
DDoS	41835
DoS GoldenEye	10293
FTP-Patator	7938
SSH-Patator	5897
DoS slowloris	5796
DoS Slowhttptest	5499
Bot	1966
Web Attack – Brute Force	1507
Web Attack – XSS	652

Partizionamento del dataset

a% = 40% *(normal) Training set*
b% = 20% *(normal) Validation set*
c% = 20% *(normal+attack) Threshold set*
d% = 20% *(normal+attack) Testing set*

https://www.unb.ca/cic/datasets/ids-2017.html

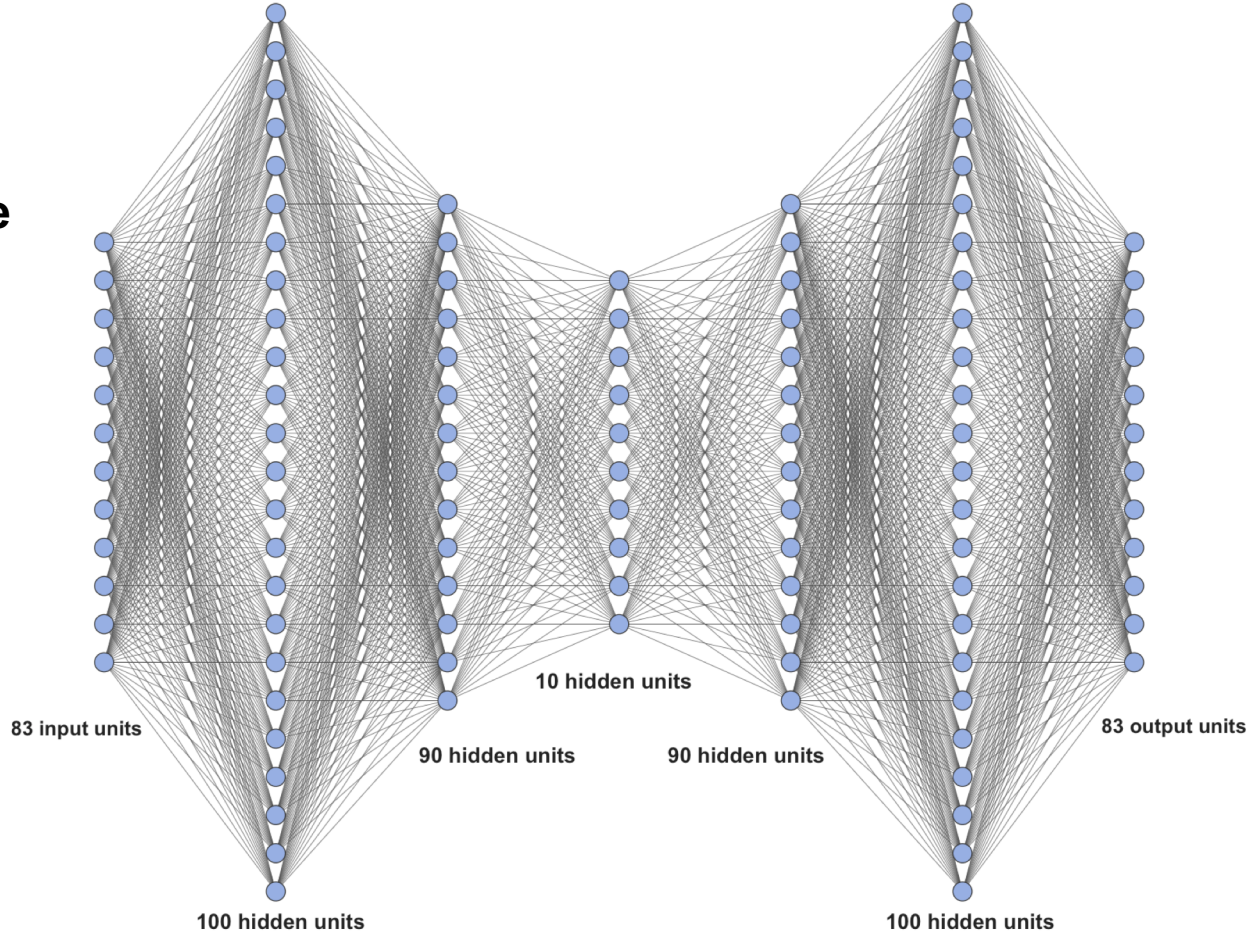
Ambiente



Sperimentazione (2)

Tuning della rete

Batch size = 200
Epoche = 100



Risultati Sperimentali

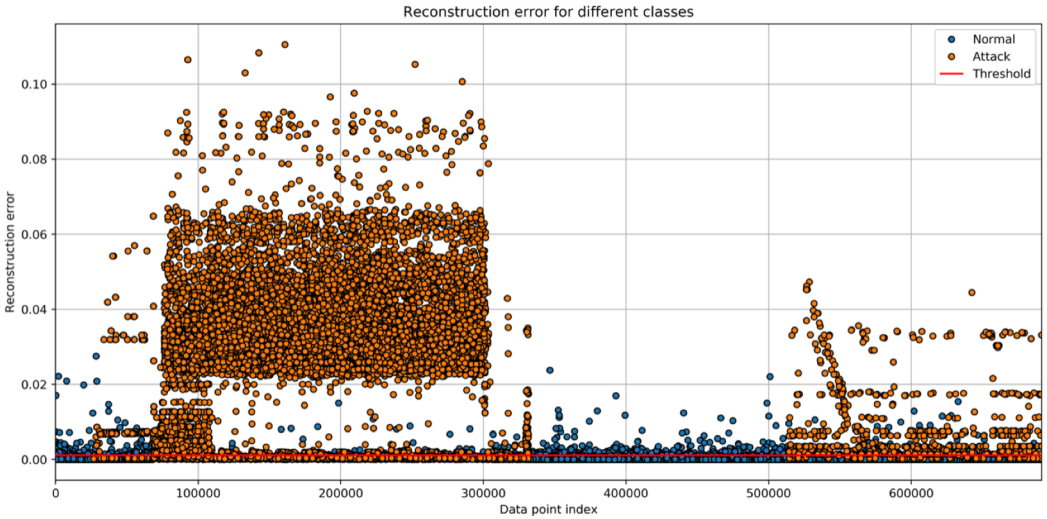
Metriche

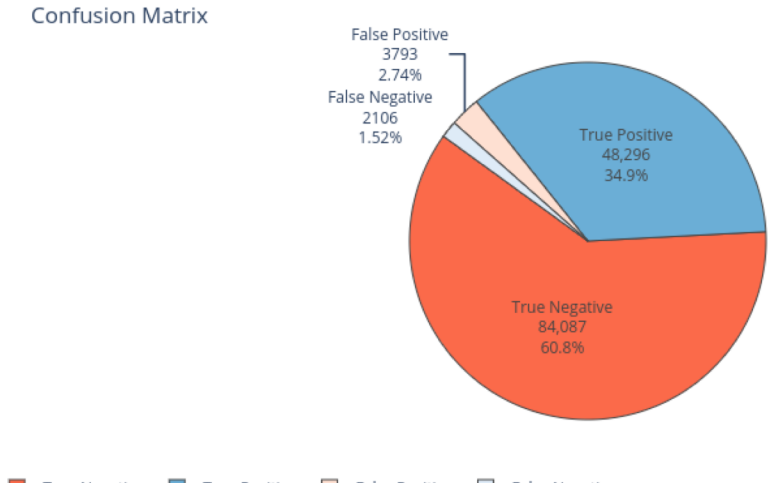
$$Detection\ Rate = \frac{TP}{(TP+FN)}$$

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)}$$

$$False\ Alarm\ rate = \frac{FP}{(FP+TN)}$$

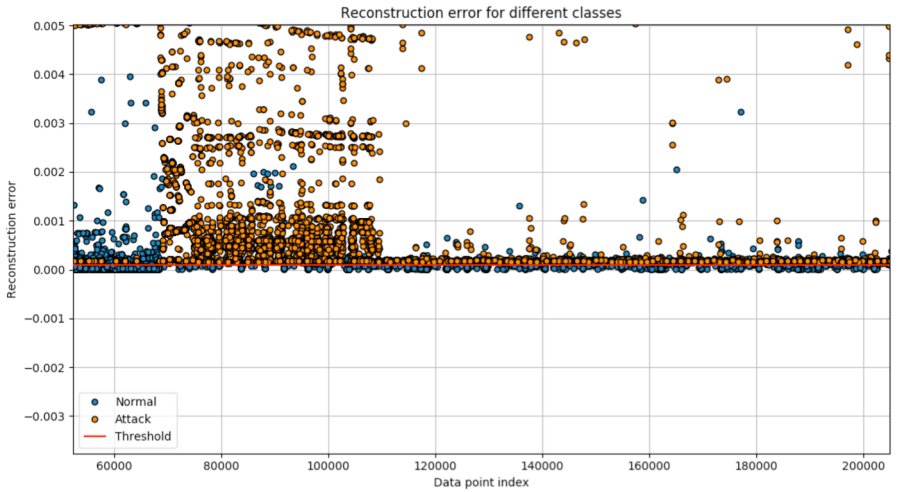
DoS



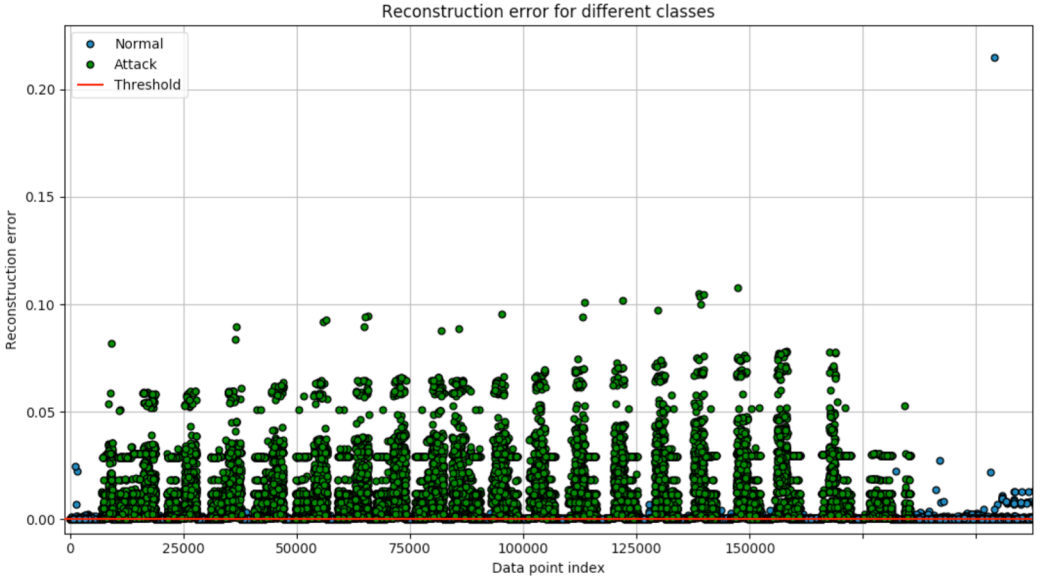


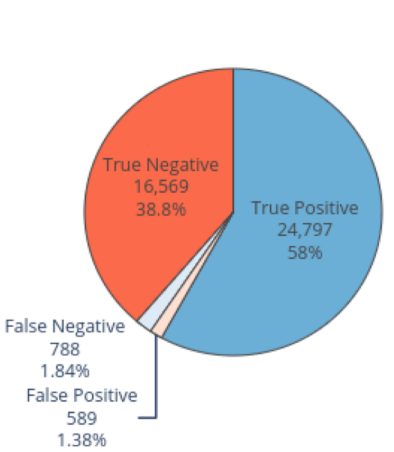
Detection Rate %	95.82
Accuracy %	95.73
False alarm rate %	4.32

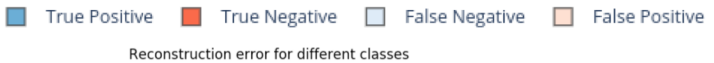
zoomed plot →



DDoS

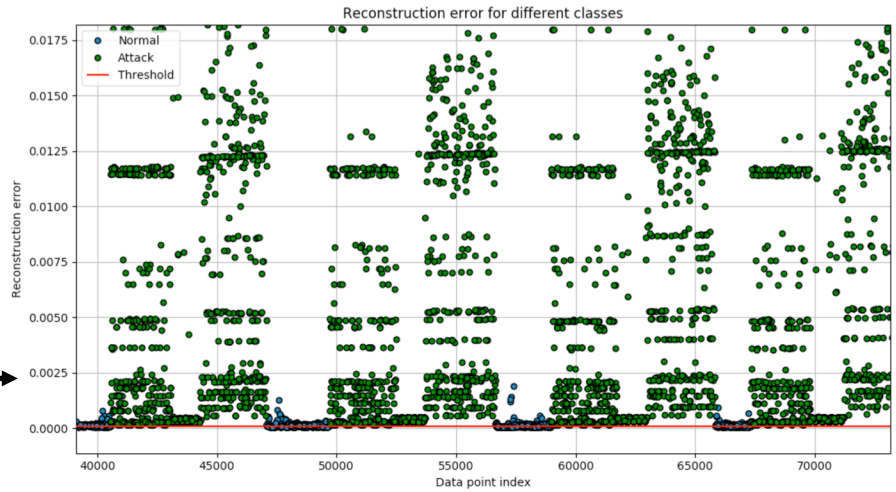




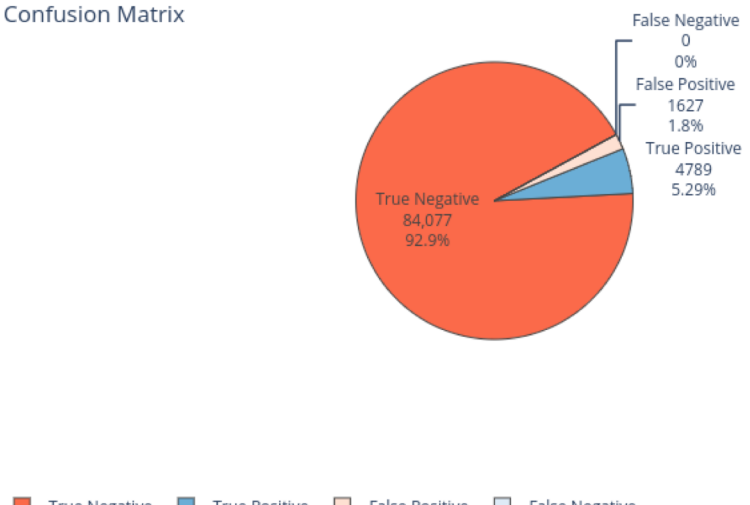


Detection Rate %	96.92
Accuracy %	96.78
False alarm rate %	3.43

zoomed plot →

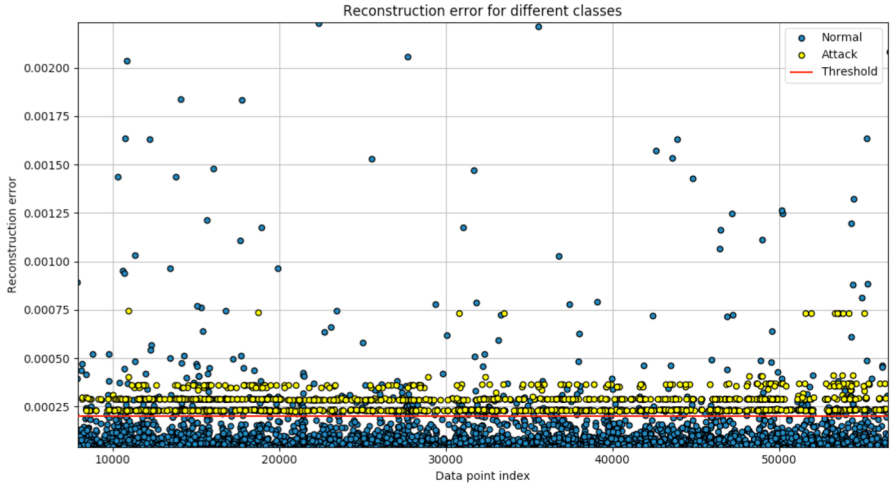


Brute Force

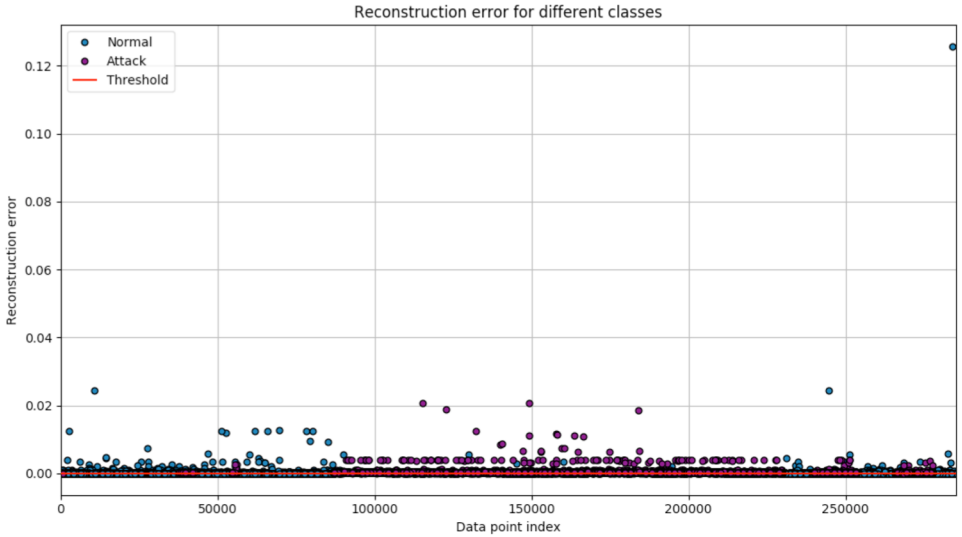


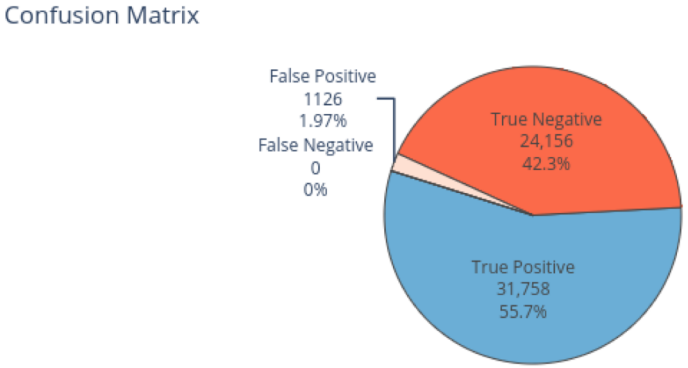
Detection Rate %	100
Accuracy %	98.20
False alarm rate %	1.9

zoomed plot →



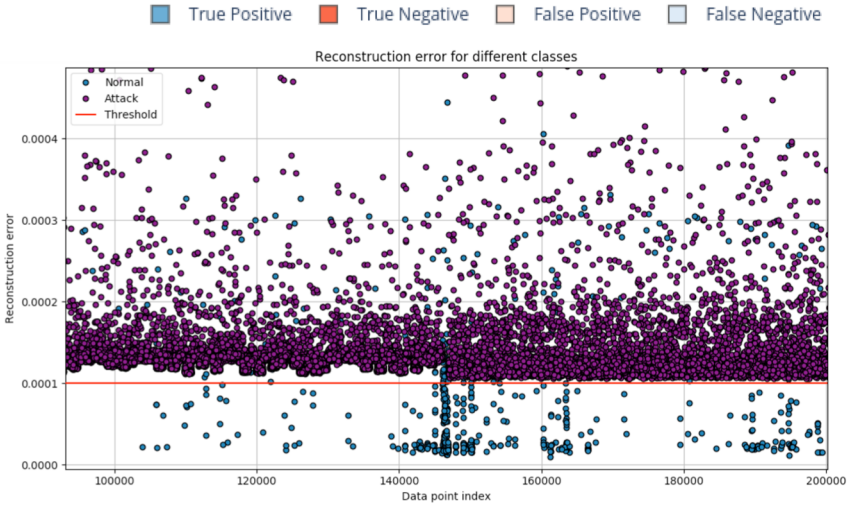
Port Scan



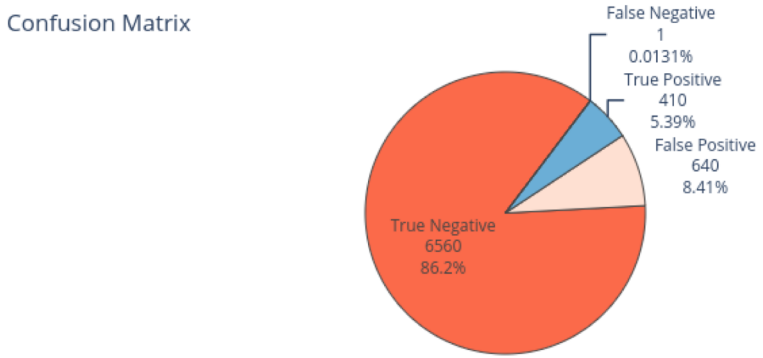


Detection Rate %	100
Accuracy %	98.03
False alarm rate %	4.45

zoomed plot →

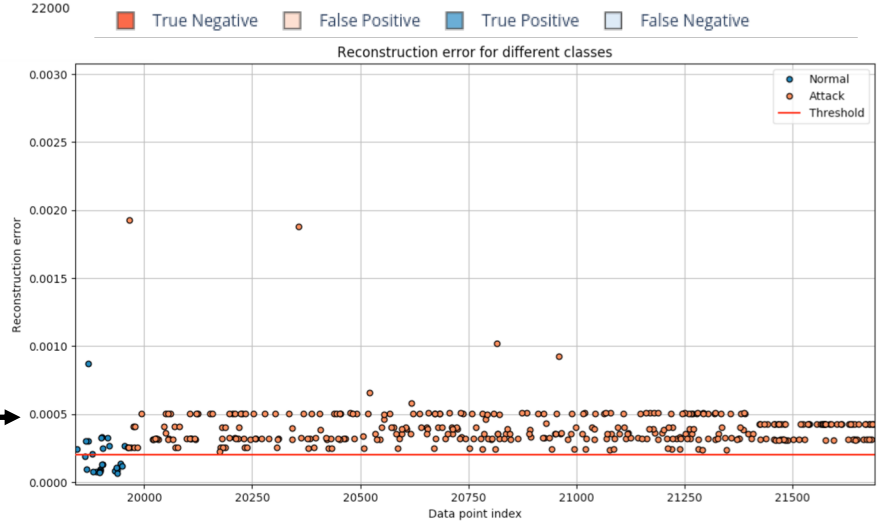


Bot

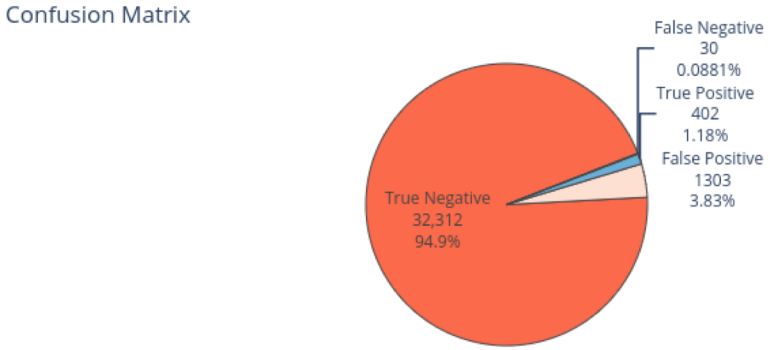


Detection Rate %	99.76
Accuracy %	91.58
False alarm rate %	8.89

zoomed plot →



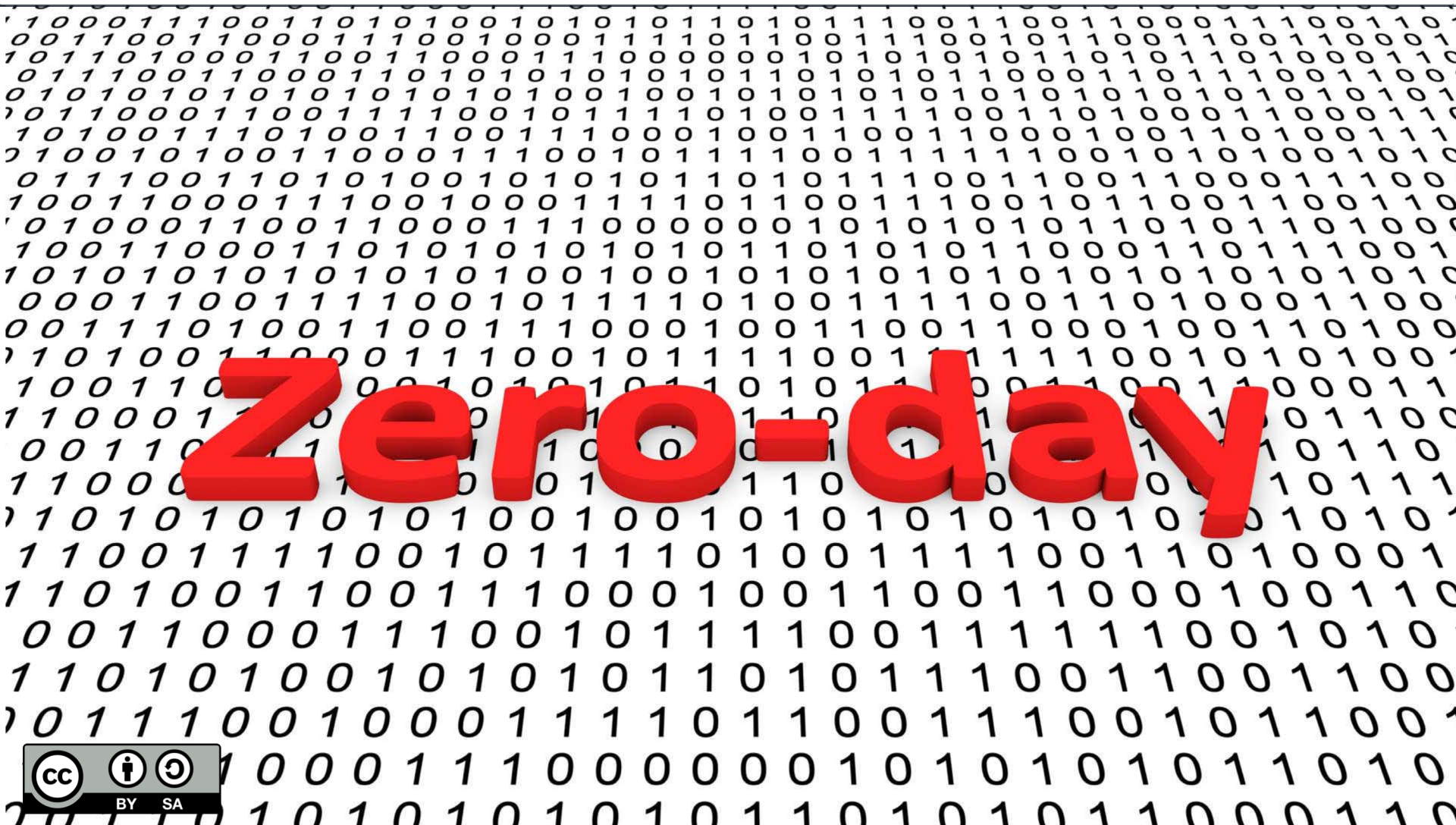
Web Attack



Detection Rate %	93.06
Accuracy %	96.08
False alarm rate %	3.88

zoomed plot →

0-day discovery



0-day discovery (1)

Obiettivo: riconoscere attacchi mai visti prima

Test sui DoS

Traffic	Number of instances
Dos Hulk	231073
DoS GoldenEye	10293
DoS SlowHTTPTest	5499
DoS Slowloris	5796
Heartbleed	11
Benign	440031

Hulk trattato come uno 0-day

- Training del modello senza Hulk
- Comparazione con classificatori supervised

Scenario 1 – Hulk known

Learning set

- DoS Hulk
- DoS GoldenEye
- DoS Slowloris
- DoS SlowHTTPTest
- Heartbleed

Test set

- DoS Hulk
- DoS GoldenEye
- DoS Slowloris
- DoS SlowHTTPTest
- Heartbleed

Scenario 2 – Hulk 0day

Learning set

- DoS GoldenEye
- DoS Slowloris
- DoS SlowHTTPTest
- Heartbleed

Test set

- DoS Hulk
- DoS GoldenEye
- DoS Slowloris
- DoS SlowHTTPTest
- Heartbleed

0-day discovery (2)

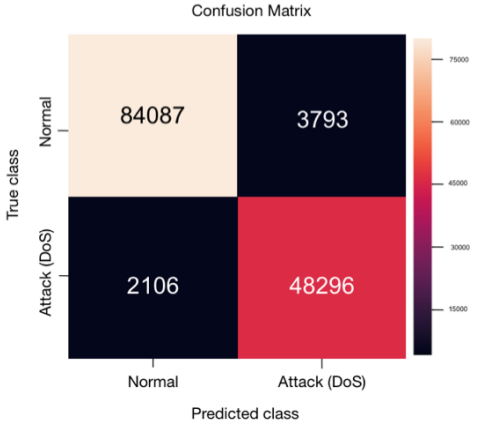
Comparazione delle performance

Confronto con classificatori supervised

- Random Forest
- QDA
- Naive Bayes

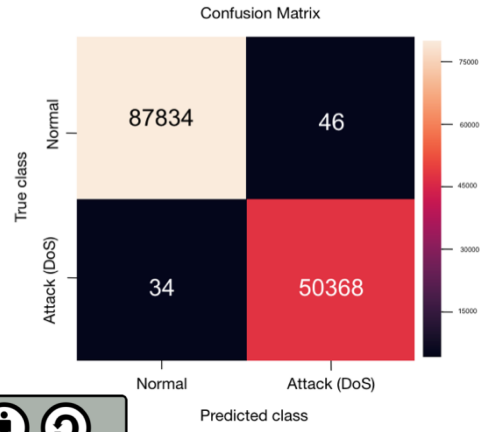
0-day discovery (3)

Matrici di confusione – Hulk-known

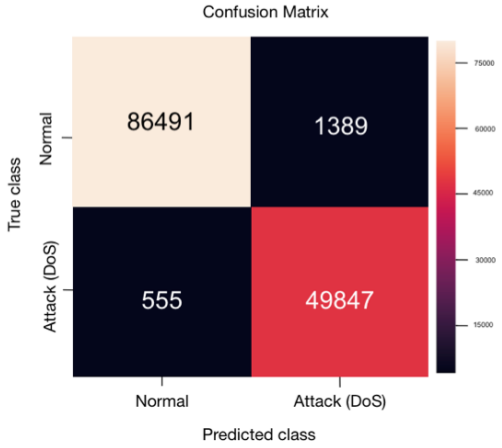


→ Modello proposto

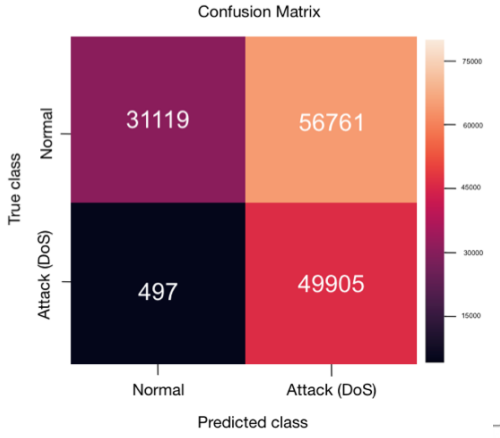
Random Forest



QDA

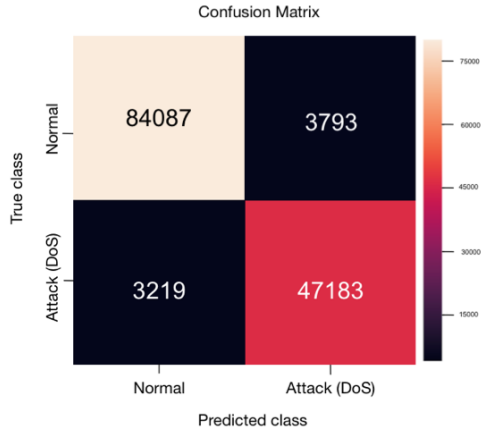


Naive Bayes



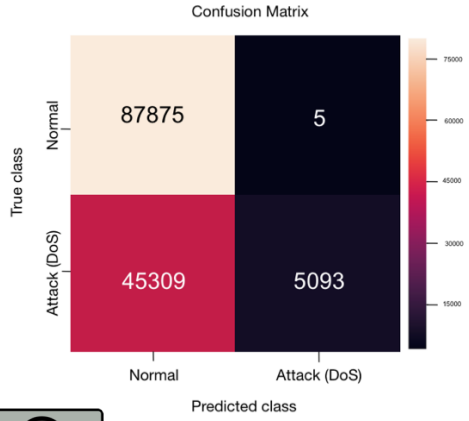
0-day discovery (4)

Matrici di confusione – Hulk-0day

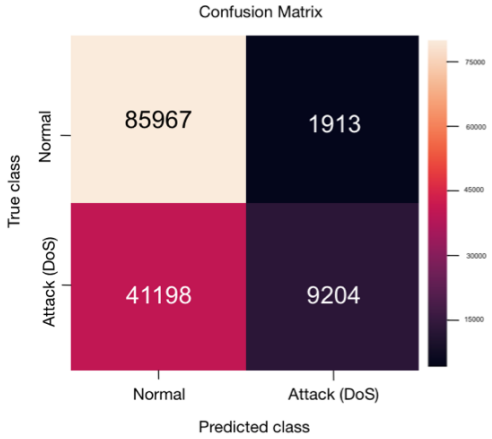


→ Modello proposto

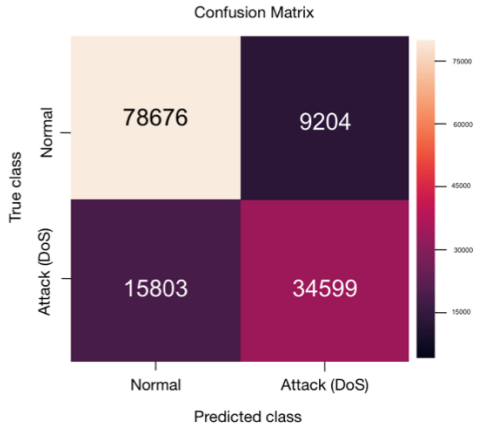
Random Forest



QDA



Naive Bayes



0-day discovery (5)

Comparazione delle performance
metriche

Network/algorithm	Detection rate %		Accuracy %		Precision %		False alarm rate %	
	Hulk-known	*Hulk-0day*	*Hulk-known*	*Hulk-0day*	*Hulk-known*	*Hulk-0day*	*Hulk-known*	*Hulk-0day*
Random Forest	99.93	10.00	99.94	67.23	99.91	99.90	0.05	0.01
QDA	98.90	18.26	98.59	68.82	97.29	82.79	1.58	2.18
Naive Bayes	99.01	68.65	58.59	81.92	46.79	78.99	64.59	10.47
ZED-IDS AE	95.82	93.61	95.73	94.93	92.72	92.56	4.32	4.32

modello proposto

Conclusioni

- Accuratezza del 95.73% per lo scenario 2
  - Modello potenzialmente utile per il riconoscimento di 0-day
- Tempi di training e di detection bassi
- Possibile integrazione in tool per la detection di attacchi real-time

«Change is challenging. And security is like a moving target, so make sure you are able to deal with and work through frequent changes.».

— Cindi Carter

Marta Catillo
martacatillo@gmail.com